

An Introduction to Parallel Supercomputing

Pete Beckman

Argonne National Laboratory





MCS Division meeting c. 1983

- “If our R&D is going to be relevant ten years from now, we need to shift our attention to parallel computer architectures”
- “Los Alamos has a Denelcor HEP: let’s experiment with it”



POOMA Project: 1996

John Reynders

Parallel Platform Paradox

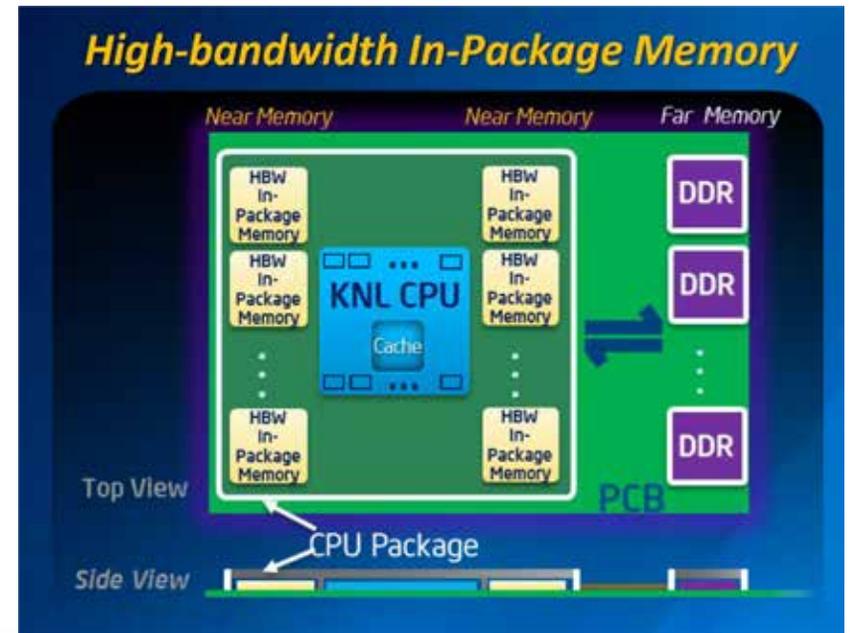
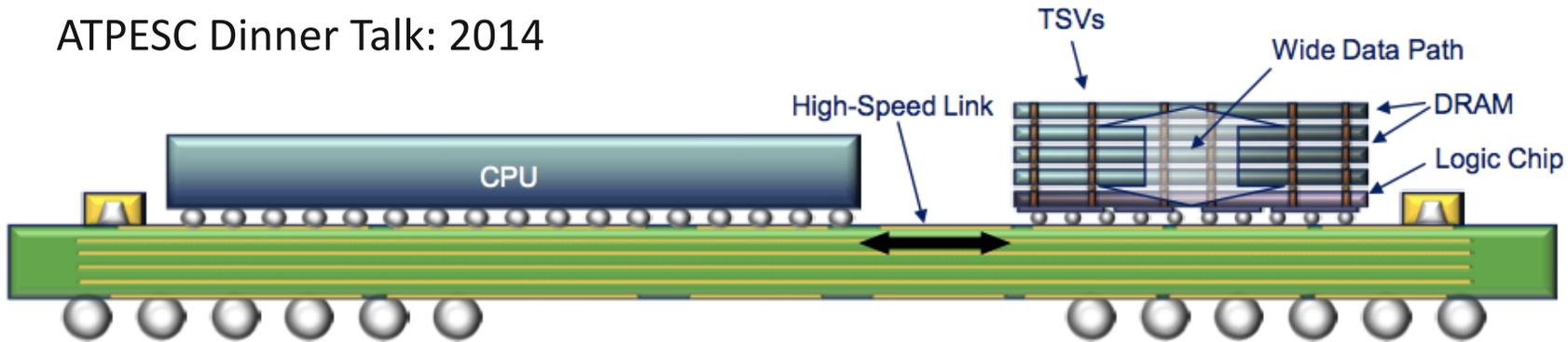
“The average time required to implement a moderate-sized application on a parallel computer architecture is equivalent to the half-life of the latest parallel supercomputer.”



“Although a strict definition of “half-life” could be argued, no computational physicist in the fusion community would dispute the fact that most of the time spent implementing parallel simulations was focused on code maintenance, rather than on exploring new physics. Architectures, software environments, and parallel languages came and went, leaving the investment in the new physics code buried with the demise of the latest supercomputer. There had to be a way to preserve that investment.”

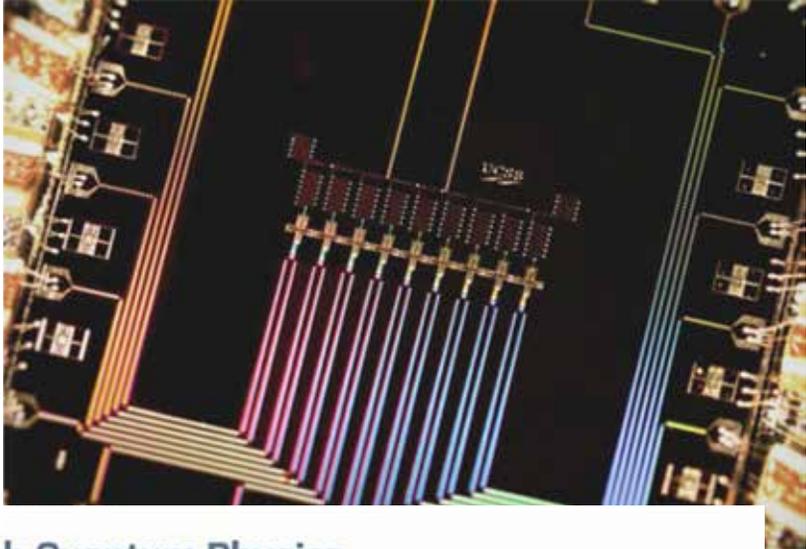
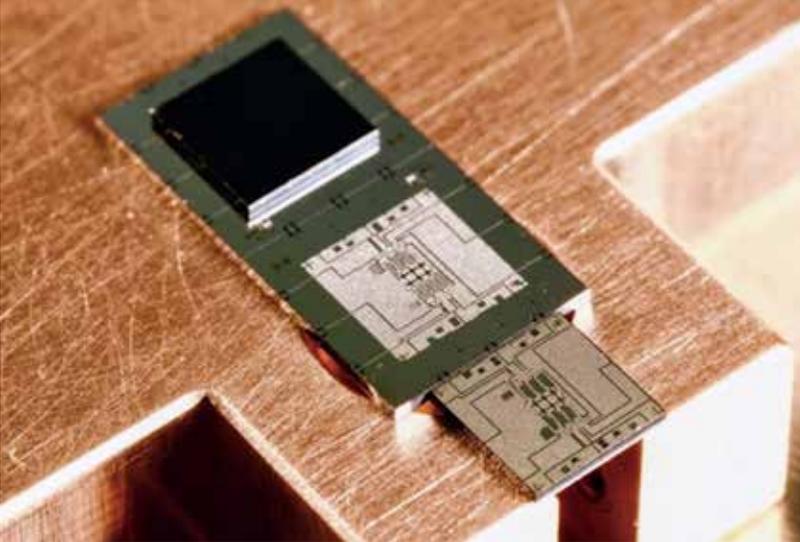


ATPESC Dinner Talk: 2014



Today...

Quantum?



Computing with Quantum Physics
A faster, cheaper path to exascale

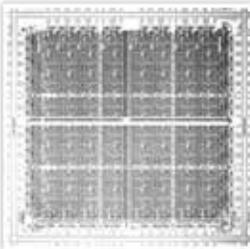
QUANTUM CHIP DESIGN CHIP MICROFABRICATION ADVANCED CONTROL ELECTRONICS CLOUD SOFTWARE & APPLICATIONS

Pete Beckman Argonne National Laboratory



INTEL MARRYING FPGA, BEEFY BROADWELL FOR OPEN COMPUTE FUTURE

March 14, 2016 Nicole Hemsoth



For those who read here often, there are clear signs that the FPGA is set to become a **compelling acceleration story** over the next few years.

From the relatively recent Intel acquisition of Altera by chip giant Intel, to less talked-about advancements on the programming front (OpenCL progress, advancements in both hardware and software from FPGA competitor to Intel/Altera, Xilinx) and of course, consistent competition for the compute acceleration market from GPUs, which dominate the coprocessor market for now

Last week at the Open Compute Summit we finally got a glimpse of one of the mini-server FPGAs that might fit into announcements about the new Xeon chip, the Broadwell E, and a diagram that

ADOPTION OF INTEL FGAs FOR ACCELERATION OF ENTERPRISE WORKLOADS GOES MAINSTREAM



FPGA....

Microsoft Catapult



- Two 8-core Xeon 2.1 GHz CPUs
- 64 GB DRAM
- 4 HDDs @ 2 TB, 2 SSDs @ 512 GB
- 10 Gb ethernet
- No cable attachments to server

Air flow
200 LFM
68°C Inlet

FPGAs:

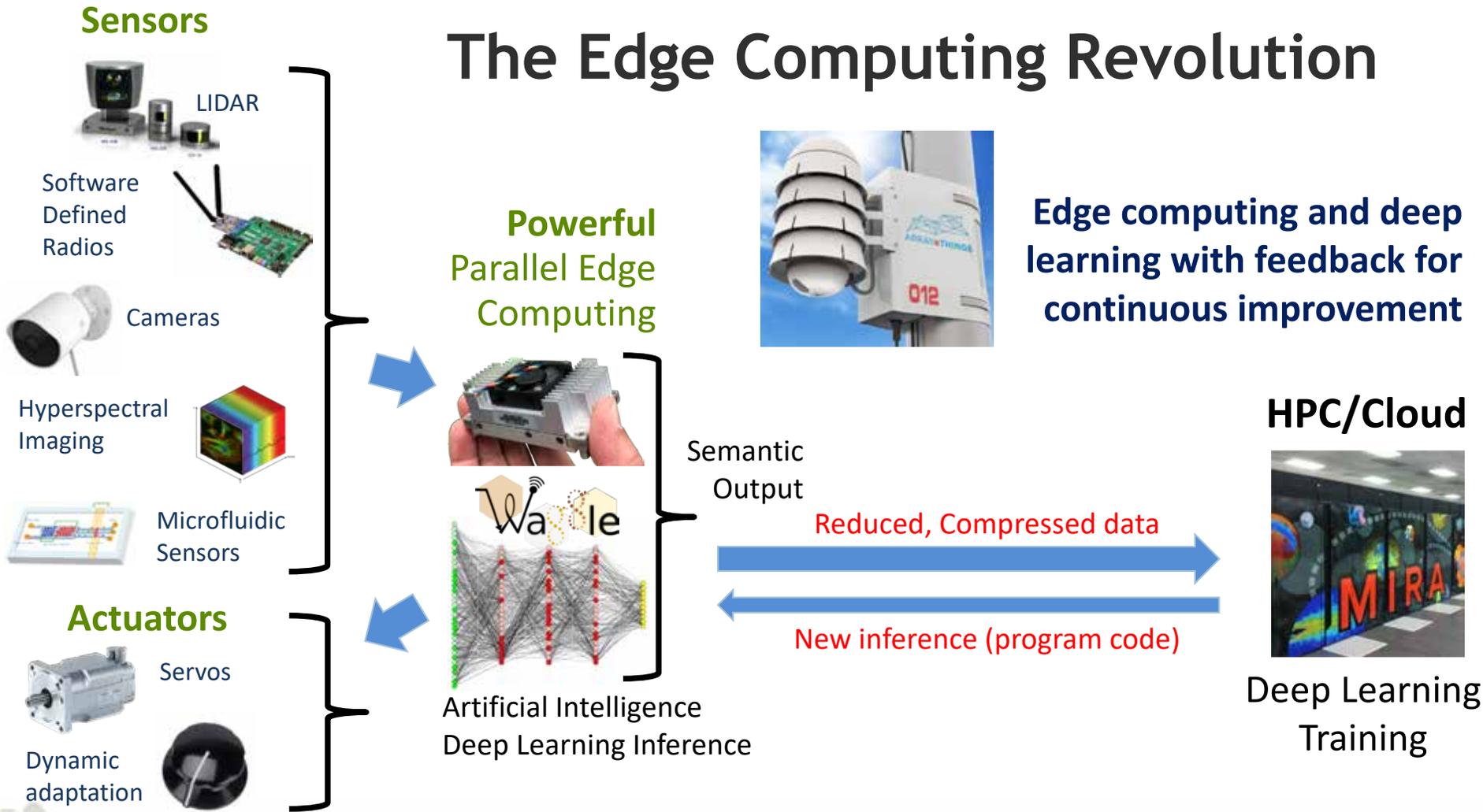
- Biggest win: SDN goes into FPGA/NIC
- All servers since 2015 deploy this
- Cloud providers recognize we need "fluid" systems, but HPC lags behind...



Andrew Putnam



The Edge Computing Revolution



Edge computing and deep learning with feedback for continuous improvement



Google TensorFlow Processors

Processor Designed for Deep Learning fujitsu

Utilizing technologies derived from the K-computer

DLU™ (Deep Learning Unit)

FY2018 ~

Features of DLU

- Architecture designed for Deep Learning
- Low power consumption design
- Optimized precision
- Goal: 10x Performance / Watt compared to competitors
- Scalable design with Tofu interconnect technology
- Ability to handle large-scale neural networks



GRAPHCORE

groq

MYTHIC



Intel Myriad



NVIDIA TX2



Amazon DeepLens

10 minutes to your first deep learning project

- 1 Choose your deep learning model from the AWS DeepLens pre-trained model library, or your own models trained with Amazon SageMaker.
- 2 Deploy your model to the device with a single click.
- 3 Watch the results in real time in the AWS Management Console.



Devices (1)

Search devices

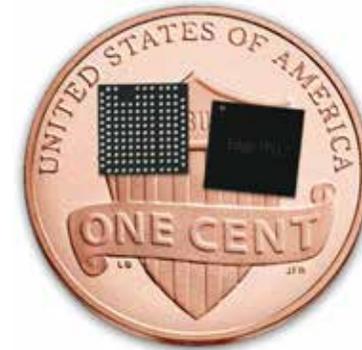
Name	Project	Registration status
Pete-Hacking	Face-detection	Registered

Projects (1)

Deploy to device Actions Create new project

Search projects

Name	Description	Version	Creation time	Last updated
Face-detection	Detect all faces in your surroundings	1	6/18/2018, 4:39:04 PM	6/26/2018, 3:52:25 AM



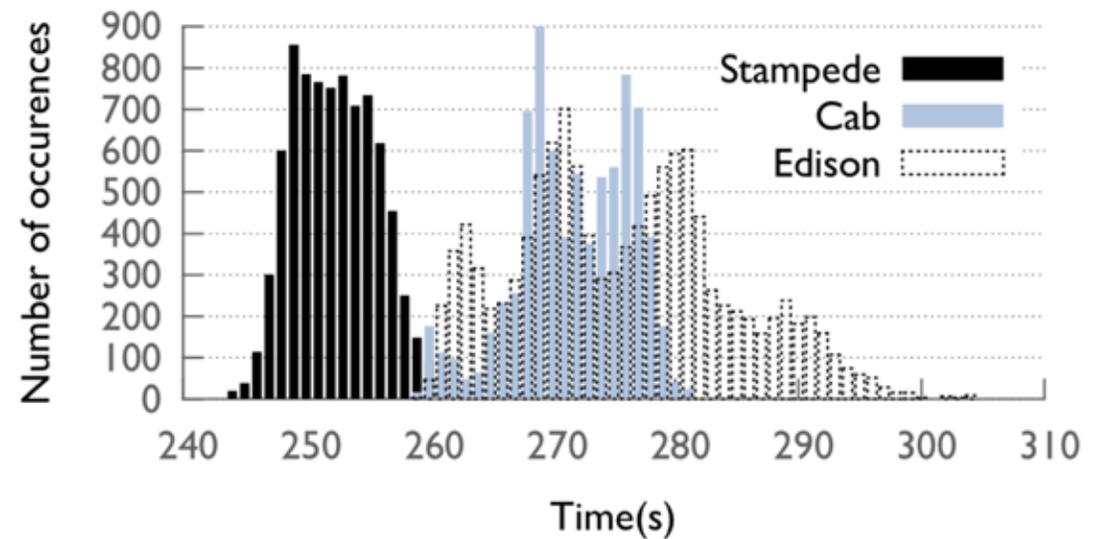
Google Edge TPU

July 2018

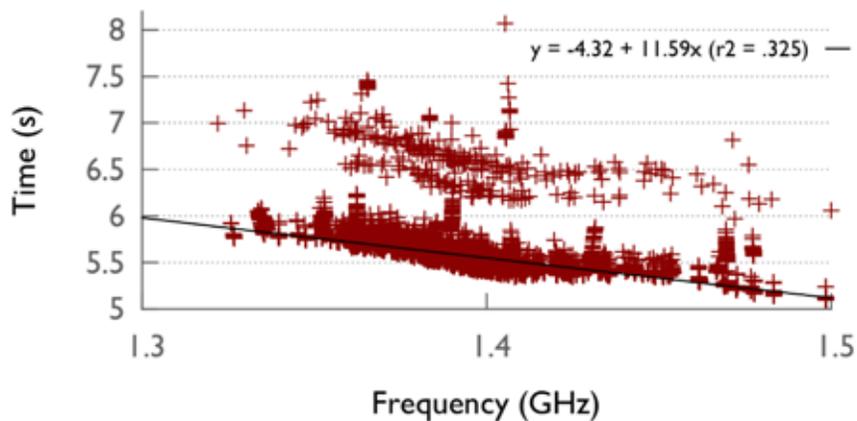
“Edge TPUs are designed to complement our Cloud TPU offering, so you can accelerate ML training in the cloud, then have lightning-fast ML inference at the edge. Your sensors become more than data collectors — they make local, real-time, intelligent decisions.”

Equal Work is not Equal Time

Histogram of Execution Time



KNL Frequency and Execution Time Variability



Courtesy Bilge Acun: Dissertation, UIUC, 2018



- **OPM (Other People's Math (libraries))**
- **Encapsulation**
 - Parallelism & Messaging & I/O
- **Embedded Capabilities**
 - Debugging
 - Performance Monitoring
 - Correctness Detection
 - Resilience
- **The Two Workflow Views**
 - Science: (problem setup, analysis, etc.)
 - Programmer: (mod, testing, document, commit)
- **Automation**
 - A+ Build system, nightly test and build, configuration
 - Embedded versioning and metadata
- **Community: web, tutorial, email, bug tracking, etc**

Pete's Investment Recommendations





Memory Heterogeneity Variability



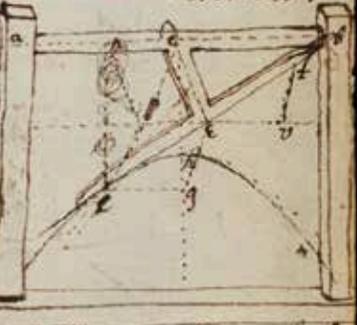
Of Refractions.

1 If y^e ray ac be refracted at the center of y^e circle acy towards d at abt be y^e center. Then suppose $ab : Ed :: d : e$. See Cartes Dioptricks

2 If there be an hyperbola a b whose distance of whose foci are to its transverse axis hf as d to e . Then y^e ray ac is refracted to y^e exterior focus d. See C. Dioptricks

3 Having y^e proportion of d to e , or bd to hf . The Hyperbola may be thus described.

1 Upon y^e center a. b let y^e instrument adbec be moved up with instrum observe y^e adt de t c fct of y^e the beams cut is not in y^e same plane at adbt but intersects it at y^e angle tsv soe y^e t to t sv. then $d : e :: ct : tv$. Or $de :: Rad : \sin$ of $\angle tv$. also make $de = \frac{1}{2}$ i.e. half y^e transverse diameter



This piece of plate chm in the same plane wth ab. If moving y^e instrument adbec to ac fro its edge cut shall cut or wear it into y^e shape of y^e desired Parabola. Or the plate chm may be filed away until y^e edge cut exactly touch it every where.

2 By the same proceeding Des-Cartes concave Hyperbolicall wheels may be described by being turned wth a chissell the whose edge is a straight line inclined to the axis of the manevill by y^e angle near angle is found by making $d : e :: ct : tv :: Rad : \sin$ of $\angle tv$

3 By the same reason a wheel may be turned Hyperbolicallly concave of Hyperbola being convex. Or a Plate may be turned Hyperbolicallly concave